

WHAT IS CLAIMED IS:

1 1. A method for loading input data in one or more hierarchical format input
2 files into a data store, comprising:
3 performing parallel processing of one or more input files to output data; and
4 serially loading the data into the data store while enforcing the order of the data in
5 the one or more input files.

1 2. The method of claim 1, further comprising:
2 receiving a physical input file; and
3 logically dividing the physical input file into multiple sections, wherein each of the
4 multiple sections is an input file.

1 3. The method of claim 2, further comprising:
2 while performing processing of a first section from the multiple sections under
3 control of a first row mapper,
4 determining that there has been an error in logically dividing the physical
5 input file;
6 continuing processing of a next section from the multiple sections that is
7 also being processed by a second row mapper; and
8 notifying the second row mapper to terminate processing of the next
9 section.

1 4. The method of claim 3, wherein the data from each of the input files is
2 appended to a separate temporary storage location and further comprising:
3 deleting the temporary storage location into which the second row mapper was
4 appending the data from the processing of the next section.

1 5. The method of claim 1, wherein serially loading the data further comprises:
2 loading the data without generating SQL commands.

1 6. The method of claim 1, wherein the data from each of the input files is
2 appended to a separate temporary storage location and further comprising:
3 when serial loading is interrupted, restarting the serial loading using the data in the
4 separate temporary storage locations without reprocessing the one or more input files.

1 7. The method of claim 1, wherein the parallel processing is performed by two
2 or more row mappers.

1 8. A method for loading input data in one or more hierarchical format input
2 files into a data store, comprising:
3 under control of a master row mapper,
4 invoking one or more slave row mappers, wherein the slave row mappers
5 perform processing in parallel with the master row mapper and with each other;
6 processing data in a first input file; and
7 serially loading the processed data and data in one or more spillfiles into
8 the data store; and
9 under control of each of the slave row mappers,
10 processing data in a separate input file; and
11 storing results of the processing in a corresponding spillfile.

1 9. The method of claim 8, further comprising:

2 under control of the master row mapper,
3 determining that there has been an error in processing the data in at least
4 one input file; and
5 terminating the slave row mappers.

1 10. The method of claim 8, further comprising:
2 under control of the master row mapper,
3 determining that there has been an error in loading the processed data in at
4 least one input file; and
5 terminating the slave row mappers.

1 11. The method of claim 8, further comprising:
2 under control of at least one of the slave row mappers,
3 determining that there has been an error in processing the data in at least
4 one input file; and
5 terminating each of the other slave row mappers processing a separate
6 input file whose order follows the separate input file being processed by the slave row
7 mapper that determined that there has been an error.

1 12. The method of claim 8, wherein each of the one or more input files is a
2 section, further comprising:
3 under control of the master row mapper and each of the slave row mappers, during
4 processing of a current section, at the end of each processing unit,
5 determining that processing has crossed into a next section; and
6 continuing to process data in the next section.

1 13. The method of claim 8, further comprising:

2 when restarting loading of the processed data, skipping a specified number of rows
3 in at least one of the input files.

1 14. An article of manufacture including a program for loading input data in one
2 or more hierarchical format input files into a data store, wherein the program causes
3 operations to be performed, the operations comprising:
4 performing parallel processing of one or more input files to output data; and
5 serially loading the data into the data store while enforcing the order of the data in
6 the one or more input files.

1 15. The article of manufacture of claim 14, wherein the operations further
2 comprise:
3 receiving a physical input file; and
4 logically dividing the physical input file into multiple sections, wherein each of the
5 multiple sections is an input file.

1 16. The article of manufacture of claim 15, wherein the operations further
2 comprise:
3 while performing processing of a first section from the multiple sections under
4 control of a first row mapper,
5 determining that there has been an error in logically dividing the physical
6 input file;
7 continuing processing of a next section from the multiple sections that is
8 also being processed by a second row mapper; and
9 notifying the second row mapper to terminate processing of the next
10 section.

1 17. The article of manufacture of claim 16, wherein the data from each of the
2 input files is appended to a separate temporary storage location and wherein the
3 operations further comprise:

4 deleting the temporary storage location into which the second row mapper was
5 appending the data from the processing of the next section.

1 18. The article of manufacture of claim 14, wherein the operations for serially
2 loading the data further comprise:

3 loading the data without generating SQL commands.

1 19. The article of manufacture of claim 14, wherein the data from each of the
2 input files is appended to a separate temporary storage location and wherein the
3 operations further comprise:

4 when serial loading is interrupted, restarting the serial loading using the data in the
5 separate temporary storage locations without reprocessing the one or more input files.

1 20. The article of manufacture of claim 14, wherein the parallel processing is
2 performed by two or more row mappers.

1 21. An article of manufacture including a program for loading input data in one
2 or more hierarchical format input files into a data store, wherein the program causes
3 operations to be performed, the operations comprising:

4 under control of a master row mapper,

5 invoking one or more slave row mappers, wherein the slave row mappers
6 perform processing in parallel with the master row mapper and with each other;

7 processing data in a first input file; and

8 serially loading the processed data and data in one or more spillfiles into
9 the data store; and

10 under control of each of the slave row mappers,
11 processing data in a separate input file; and
12 storing results of the processing in a corresponding spillfile.

1 22. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 under control of the master row mapper,
4 determining that there has been an error in processing the data in at least
5 one input file; and
6 terminating the slave row mappers.

1 23. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 under control of the master row mapper,
4 determining that there has been an error in loading the processed data in at
5 least one input file; and
6 terminating the slave row mappers.

1 24. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 under control of at least one of the slave row mappers,
4 determining that there has been an error in processing the data in at least
5 one input file; and
6 terminating each of the other slave row mappers processing a separate
7 input file whose order follows the separate input file being processed by the slave row
8 mapper that determined that there has been an error.

1 25. The article of manufacture of claim 21, wherein each of the one or more
2 input files is a section and wherein the operations further comprise:
3 under control of the master row mapper and each of the slave row mappers, during
4 processing of a current section, at the end of each processing unit,
5 determining that processing has crossed into a next section; and
6 continuing to process data in the next section.

1 26. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 when restarting loading of the processed data, skipping a specified number of rows
4 in at least one of the input files.

1 27. A computer system having at least one program for loading input data in
2 one or more hierarchical format input files into a data store, comprising:
3 performing parallel processing of one or more input files to output data; and
4 serially loading the data into the data store while enforcing the order of the data in
5 the one or more input files.

1 28. The computer system of claim 27, further comprising:
2 receiving a physical input file; and
3 logically dividing the physical input file into multiple sections, wherein each of the
4 multiple sections is an input file.

1 29. The computer system of claim 28, further comprising:
2 while performing processing of a first section from the multiple sections under
3 control of a first row mapper,
4 determining that there has been an error in logically dividing the physical
5 input file;

6 continuing processing of a next section from the multiple sections that is
7 also being processed by a second row mapper; and
8 notifying the second row mapper to terminate processing of the next
9 section.

1 30. The computer system of claim 29, wherein the data from each of the input
2 files is appended to a separate temporary storage location and further comprising:
3 deleting the temporary storage location into which the second row mapper was
4 appending the data from the processing of the next section.